

빅데이터 분석방법을 활용한 제조업 혁신성과예측 방법에 대한 연구 : 딥 러닝 알고리즘을 중심으로[†]

Forecasting Innovation Performance via Deep Learning Algorithm :
A Case of Korean Manufacturing Industry

황정재(Hwang Jeong-jae)*, 김재영(Kim Jae Young)**, 박재민(Park Jaemin)***

목 차

- | | |
|-------------|-------------|
| I. 서 론 | IV. 실증분석 결과 |
| II. 선행연구 분석 | V. 결론 및 시사점 |
| III. 연구설계 | |

국 문 요 약

기술혁신에는 본질적인 어려움이 따르는데, 이는 상당부분 기술이 지닌 불확실성에 기인한다. 따라서 혁신과정에서 불확실성에 따른 위험을 감소시키기 위한 예측 방법론은 정량적 분야와 정성적 분야 모두에서 제시되어 왔다. 한편 최근 빅 데이터와 인공지능에 큰 관심이 이어지며 특히 알파고의 알고리즘 중 하나인 딥 러닝이 뛰어난 성능을 보이고 있다. 이에 본 연구는 혁신성과 예측에 있어 딥 러닝을 이용한 방법론을 접목하여 연구를 진행하였다. 모델 구축 및 학습에 있어 KIS 2016 데이터를 이용하였으며, 투입 요인으로는 정보 원천의 사용도와 혁신 목적을 사용하였고 산출 요인으로는 혁신 성과 지표를 구성하여 사용하였다. 분석 결과 선행 연구들에 비해 예측의 정확도가 향상되었음을 확인할 수 있었다. 또한 학습이 진행됨에 따라 예측의 자유도 역시 향상됨을 확인하였다.

핵심어 : 혁신예측, 빅 데이터, 한국기업혁신조사, 딥 러닝, 머신러닝

※ 논문접수일: 2017.12.13, 1차수정일: 2018.4.21, 게재확정일: 2018.5.10

* 건국대학교 기술경영학과 석사과정, ohnhop@naver.com, 주저자

** 건국대학교 기술경영학과 석사과정, miracle013@naver.com

*** 건국대학교 기술경영학과 교수, jpark@konkuk.ac.kr, 교신저자

† 본 논문은 '2017 한국기술혁신학회 추계학술대회(2017. 11. 2)'에서 발표되어 우수논문으로 선정된 초고를 수정·보완하여 작성하였음.

ABSTRACT

Technological innovation has inherent difficulties, largely due to the uncertainties of technology. Thus, the forecasting methodology to reduce the risk of uncertainty in the innovation process has been presented both in quantitative and qualitative fields. On the other hand, big data and artificial intelligence have attracted great interest recently, and deep learning, which is one of the algorithms of AlphaGo, is showing excellent performance. In this study, deep learning methodology was applied to the prediction of innovation performance. To make the prediction model, we used KIS 2016 data. The input factors were importance of information source and innovation objectives and the output factor was innovation performance index, which was calculated for this study. As a result of the analysis, it can be confirmed that the accuracy of prediction is improved compared with the previous studies. As learning progressed, the degree of freedom of prediction also improved.

Key Words : Innovation Performance, Big Data, Korea Innovatino Survey, Deep Learning, Machine Learning

I. 서론

기술은 본질적으로 동적이며 복잡하기 때문에 이에 대한 예측은 매우 어려운 일이며, 특히 기술의 동적인 측면을 강조하는 혁신의 경우 그러한 어려움이 더욱 두드러지게 된다(정선양, 2016). 따라서 이를 수행하는 데에는 상당한 시간과 자원이 필요하나 이는 충분히 들일만한 가치가 있는 일인데, 기술과 혁신을 예측하고 이에 대응하는 데에 자원을 투입할수록 불확실성에서 오는 손실을 줄여 불확실성을 계산된 위험(calculated risk)으로 만들 수 있기 때문이다(Tidd and Bessant, 2013). 이에 전문가들은 수많은 기술예측 방법을 만들어 왔으며, 이는 지금도 정량적 방법과 정성적 방법 양쪽에서 진화해 나가고 있다.

또한 혁신은 신제품의 개발뿐 아니라 생산 및 이를 통한 이윤 창출까지의 전과정을 포함하기 때문에 혁신 성과에 대한 이해 역시 중요한 연구분야로서 다루어지고 있다. 이에 혁신 성과 지표 및 성과에 영향을 미치는 요인에 대한 연구도 활발히 행해지고 있으며 주로 특허나 매출 비중과 같은 혁신 성과 지표를 설정하고 회귀모형을 활용하여 이에 R&D 활동이나 기업 규모 등의 변수들이 어떤 영향을 미치는지를 분석하는 방향으로 연구가 진행되어 왔다.

한편 최근 빅데이터와 인공지능을 필두로 한 새로운 예측 방법론들이 등장하고 있는데, 특히 2016년 3월 구글의 알파고와 이세돌 9단의 바둑대결 이후 이에 대한 관심이 점점 증대되고 있다. 특히 알파고의 알고리즘은 기존의 바둑 인공지능인 몬테카를로 트리 서치 알고리즘에 딥 러닝 알고리즘을 조합하여 구성된 것으로 알려져 있는데, 이에 딥러닝에 대한 관심은 더욱 증대되고 있다(이지훈, 2016).

이렇게 뜨거운 관심을 받고 있는 딥 러닝 알고리즘은 이미 금융 분야나 마케팅 분야에서는 널리 사용되고 있으나(이우식, 2017) 기술경영 분야에서는 타 분야에 비해 그 연구가 미진한 편이다. 이에 본 연구에서는 뛰어난 예측 능력을 지닌 것으로 알려진 딥 러닝 알고리즘을 혁신 예측에 적용하여 보고자 한다. 또한 이를 통해 불확실성을 극복하는 데에 도움을 줄 수 있는 방법론을 탐색해 보고자 한다. 이를 위하여 과학기술정책연구원에서 조사한 한국기업혁신조사 2016년 데이터(KIS 2016)을 활용하여 연구를 진행하였다.

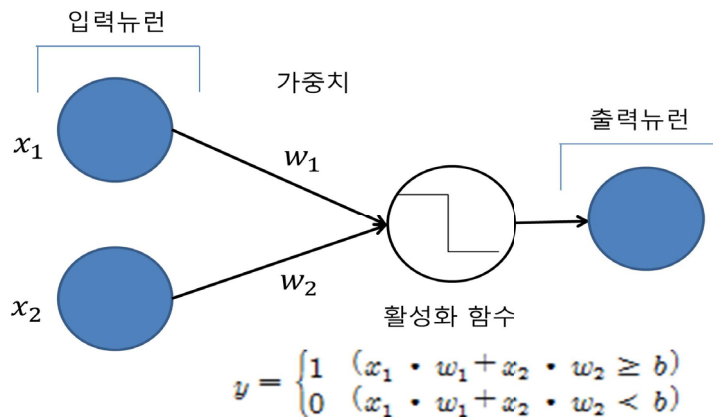
본 논문은 다음과 같이 구성된다. 2장에서는 선행연구 분석을 통하여 사회과학 분야에 딥 러닝 알고리즘이 어떻게 적용되는지를 파악할 것이며, 3장에서는 이를 바탕으로 예측 모델을 구성하고 변수를 설계한다. 그리고 4장에서는 KIS 2016 데이터를 이용한 모델 학습 및 실증분석을 진행하며 5장에서는 결론과 시사점을 도출하고자 한다.

II. 선행연구 분석

1. 딥 러닝¹⁾

1) 심층 신경망

딥 러닝은 인공신경망(artificial neural network)을 이용한 모형이다. 인공신경망 모형은 퍼셉트론(perceptron)에서 출발하며, 이는 입력층과 출력층 및 각 층을 연결하는 가중치로 구성된 구조이다(사이토 고키, 2017). 이 때 출력층은 가중치를 적용받은 입력값들이 일정한 값 이상이 되면 활성화되게 되는데, 이 때 기준이 되는 값을 편향이라 하고 이를 함수로 나타낸 것을 활성화 함수라고 한다. 이는 다음 (그림 1)과 같이 나타낼 수 있다.

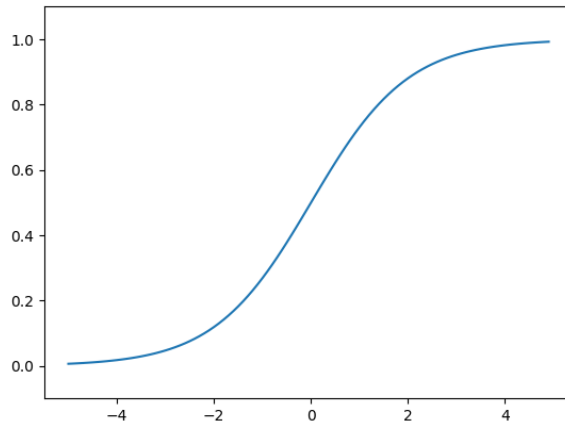


(그림 1) 퍼셉트론의 구조

퍼셉트론은 가중치와 편향을 조정하는 방법을 통해 분류 문제를 푼다. 다만 퍼셉트론은 활성화함수로 계단 모양의 함수를 사용하기 때문에 가중치의 미세한 변화가 출력값에 그대로 반영되지 않아 최적의 값을 찾는 데에 어려움이 존재한다. 그렇기 때문에 인공신경망에서는 활성화함수로 계단함수가 아닌 연속함수를 사용하게 되는데, 여기에는 sigmoid 함수나 hyperbolic tangent 함수 등이 존재한다.

이러한 연속함수의 사용으로 인해 출력값의 오차를 함수로 만들어 최솟값을 찾는 최적화 알고리즘의 사용이 가능해졌고, 이는 인공신경망이 문제를 보다 정교하게 풀 수 있게 했다.

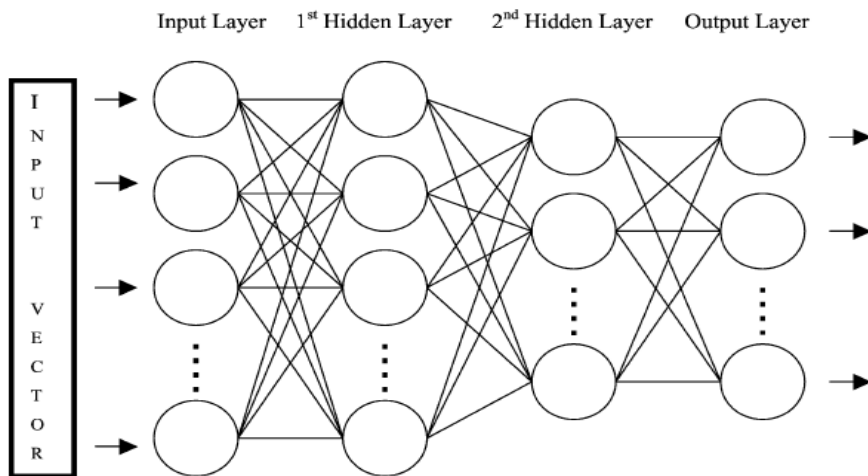
1) 익명의 심사자가 지적한 빅데이터 기법의 방법론적 특성과 관련해 연구진은 이 기법이 그 자체로 양적/질적/하이브리드 연구로 구분하기 보다는 분석의 주제에 따라 진술한 연구에 모두 적용될 수 있다고 판단하였다.



(그림 2) sigmoid 함수의 그래프

앞서 살펴본 퍼셉트론은 간단한 분류 문제들을 풀 수 있었으나 퍼셉트론이 풀 수 있는 문제는 선형 분류에 한정된다는 문제가 있었다. 따라서 비선형문제는 퍼셉트론으로만으로는 풀 수 없었으나 Minsky and Papert(1969)는 이에 대한 해결책을 제시하였다. 이는 단층으로만 되어 있는 퍼셉트론을 중첩시켜 다층으로 만드는 것이었다. 이러한 다층 구조를 통해 퍼셉트론의 표현력이 높아졌으며 인공신경망 역시 더욱 복잡한 분류 문제에 적용될 수 있게 되었다.

이렇듯 여러 층으로 구성된 인공신경망은 심층 신경망(Deep neural networks, DNN)으로 명명되었고, 이는 다음 (그림 3)과 같은 구조를 지닌다. 이 때 입력층과 출력층을 제외한 나머지는 은닉층(Hidden layer)이 된다.



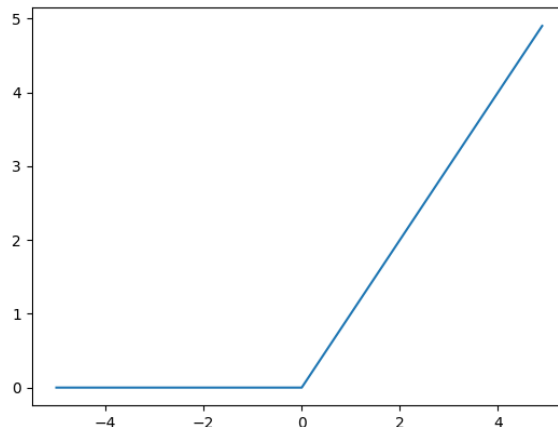
(그림 3) 심층 신경망 모형

2) 딥 러닝

이렇듯 수많은 은닉층을 지닌 심층 신경망은 오차 역전파 알고리즘을 통해 최적화된다. 오차 역전파 알고리즘은 신경망의 반대 방향으로 값을 전달하는 방식이며 이 때 전달되는 값은 노드의 편미분 값이다. 이는 가중치의 최적화에 있어 방향을 제시해주게 되는데, 이에 순전파와 역전파를 순차적으로 진행하며 가중치를 최적화하게 된다(Hopfield, 1982).

다만 심층 신경망에 오차 역전파 알고리즘을 적용하며 또 다른 한계점이 드러나게 되었는데, 이는 ‘gradient vanishing’이라는 문제였다. 가장 널리 사용되던 활성화 함수인 sigmoid 함수나 hyperbolic tangent 함수는 해당 층의 출력값을 0~1, 혹은 -1~1 사이로 제한하였는데, 결과적으로 은닉층을 많이 만들면 만들수록 전파되는 값이 작아지게 되었다. 이는 역전파시에도 동일하게 적용되어 전달되는 값이 희미해지고 입력변수들의 영향력이 제대로 전달되지 않는 문제가 생겼다.

이러한 문제는 Hinton(2006)에 의해서 해결의 실마리가 제시되었는데, gradient vanishing이 활성화 함수의 기능에 의해 발생하는 문제인 만큼 활성화 함수의 변화를 통해 이를 해결할 수 있다는 것이었다. 이에 ReLU(Rectified Linear Unit) 함수가 제시되었는데, 이는 입력값이 0보다 크면 값을 그대로 출력하고 0 이하이면 0을 출력하도록 설계되어있는 함수였다. 이는 다음 (그림 4)와 같은 그래프로 나타낼 수 있다.

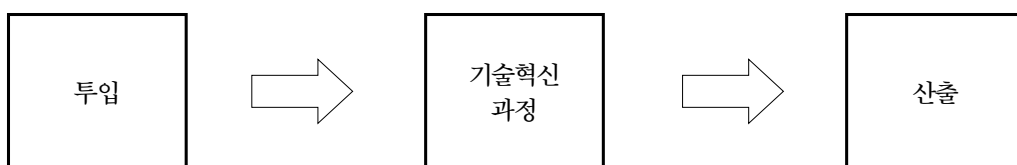


(그림 4) ReLU 함수의 그래프

이를 통해 gradient vanishing 문제는 일정부분 해결되었고, 신경망의 깊이도 보다 깊어질 수 있게 되었다. 딥 러닝은 이러한 과정을 통해 깊은 신경망을 구성하고 학습하는 모형이다.

2. 혁신 성과

기업의 혁신 성과를 측정하는 방법이나 성과에 영향을 미치는 요인에 대한 연구는 다각적으로 행해져 왔다. Tidd and Bessant(2013)은 혁신 성과의 측정방법으로 혁신 감사(innovation audit)을 제안한 바 있다. 해당 모델에서는 혁신과정을 투입, 혁신, 산출의 과정으로 보고 있는데, 혁신 감사는 이 과정 전반을 다루는 개념이다.



(그림 5) 혁신 감사에서의 혁신과정

이 중 혁신 성과와 관련이 있는 부분은 산출 부분이다. 이에 대한 대표적인 지표로는 특허권 개수, 논문의 수, 신제품의 수, 신제품의 판매/이익 비중, 고객만족도, 품질개선도, 혁신으로 인한 이득 등이 있다. 이러한 지표들을 조합하면 분석 목적에 맞는 혁신 지표를 형성할 수 있게 된다.

성태경(2005)은 KIS 2002년 데이터를 이용하여 기술혁신성과에 영향을 미치는 요인에 대한 연구를 진행한 바 있다. 해당 연구에서 혁신성과는 제품혁신이나 공정혁신이 존재하였는지 아니었는지로 설정되었으며 설명변수로는 기업 규모, 시장집중도, 협력 여부, R&D 활동 착수 여부, 기업 연령, 수출비율, 이익률, 해외지분, 산업별 연구개발집약도의 9개 변수가 활용되었다. 이에 분석 방법으로는 로지스틱 회귀분석이 사용되었으며, 분석 결과 협력 여부는 제품혁신 및 공정혁신 모두에 유의미한 영향을 미치는 것으로 나타났으며 기업 규모와 기술혁신활동, 해외지분, 이윤율의 경우 공정혁신에 대해서만 유의한 영향을 미치는 것으로 나타났다. 또한 R&D 활동은 제품혁신에만 그 영향력이 유의미하게 나타났고, 수출비율 및 기업연령의 경우에는 제품혁신과 공정혁신 모두에 그 영향력이 유의미하게 나타나지 않았다.

김성홍·김진한(2011)은 KIS 2008년 데이터를 이용하여 제품혁신 전략, 혁신 원천, 파트너의 기여도가 혁신성과에 미치는 영향에 대해 연구한 바 있다. 이 때 혁신성과로는 표준화, 시장확대, 품질원가, 유연성, 특허출원 건수, 신제품개발기간을 사용하였다. 이에 대한 설명변수는 혁신 방법이 자체개발인지 공동 협력개발인지 외부개발인지, 혁신정보의 원천이 자사인지 산업 내부인지 산업 외부인지, 파트너의 기여도가 어느 정도인지를 활용하였다. 이 때 분석방법으로

는 분산분석을 이용한 차이분석과 2단계 계층회귀분석을 사용하였으며, 세 설명변수 모두 유의미한 영향을 주는 것으로 나타났다.

천중기 외(2010)은 신발산업에서 기술혁신역량과 SCM 활동이 혁신성과에 미치는 영향에 대한 연구를 진행하였다. 이에 혁신성과로는 제품 경쟁력과 기업성장성을 사용하였으며, 연구개발역량과 기술사업화 역량은 설명변수로 사용하고 공급사슬관리 활동은 조절변수로 사용하는 회귀분석을 통해 그 영향력을 측정하였다. 해당 연구에서 설명변수들은 모두 유의한 영향을 나타내었으나 공급사슬관리 활동의 경우 연구개발 역량에는 조절효과로서 작용한 것과 달리 기업성장성에는 조절효과로서 작용하지 않은 결과가 나타났다.

이렇듯 혁신성과와 관련된 연구들에서는 연구 목적에 맞는 혁신 성과 지표를 사용하고 있었는데, 제품 경쟁력이나 매출과 관련된 변수와 특허와 관련된 변수들이 주로 사용되고 있는 것을 확인할 수 있었다. 또한 이에 대한 영향요인을 탐색하기 위한 방법으로는 로지스틱, 계층회귀분석, 조절효과 모형 등 회귀분석을 활용한 모형들이 주로 사용되고 있음을 알 수 있었다.

3. 딥 러닝 모델의 활용 사례

심층 신경망을 이용한 딥 러닝 모델은 여러 분야에서 활용되고 있으나 본 연구에서는 사회과학분야에서 해당 모델이 어떻게 사용되는지를 살펴보고 이를 기반으로 하여 혁신 성과 예측 모델을 구축하고자 한다.

사회과학 분야 중 특히 경영 분야에 있어 최근 딥 러닝이 많이 활용되어 온 것으로 보인다. 특히 2016년 이후 국내에서 이를 활용한 연구들이 많이 시행되고 있는데, 김인중 외(2010)은 딥러닝을 이용하여 T-커머스의 매출 예측을 시도하였다. 입력층에는 상품 정보, 방송 시간대, 날씨 정보 등이 포함된 623차원 벡터가 투입되었으며, 출력값은 매출액이었다. 예측에 대한 평가 지표로는 NMAE(normalized mean absolute error)가 사용되었으며, 효과적인 매출 예측을 한 것으로 평가되었다.

이외에도 주가나 환율 예측 등에 딥 러닝 알고리즘이 사용된 예도 다수 존재한다. 이지훈(2016)은 딥러닝을 활용하여 주가 예측 모델을 구축하였으며 이우식(2017)은 딥러닝 분석을 활용하여 코스피 주가지수의 방향성을 예측한 바 있다. 또한 이우식·전희주(2016)은 중국 역내·외 위안화의 변동성을 예측하는 모델을 구축한 바 있다.

다만 기술경영 분야에서는 이러한 시도가 많지 않았다. 그런데 Wang and Chein (2006)은 대만 제조업 분야를 중심으로 인공신경망 모형을 활용한 혁신 성과 예측 모델을 제시한 바 있다. 해당 연구에서는 OECD Oslo Manual에 따라 대만 제조업 기업 53개를 대상으로 설문문

진행하였고, 여기에서 투입 요인을 기술 정보 원천의 활용도와 혁신 목적으로, 산출 요인을 혁신 성과로 하여 인공신경망 모형에 적용시켰다.

Wang and Chein(2006)의 연구에서는 분석을 위하여 53개 기업 중 47개 기업의 데이터를 학습 데이터(training data)로, 나머지 6개 기업의 데이터를 테스트 데이터(test data)로 분리하였고, 학습데이터를 사용하여 모델을 학습시킨 후 테스트 데이터에 적용하여 오차를 계산하는 방법을 사용하였다. 신경망의 오차를 나타내는 지표로는 평균제곱오차(Mean Squared Error)를 사용하였으며 신경망 모델은 20개의 투입변수를 받아 14, 8개의 뉴런으로 구성된 2개의 은닉층을 통과하여 1개의 출력변수를 산출하는 모형이었다. 분석 결과 test error는 0.04~2.44 수준으로 나타났으며, 이는 예측의 가치가 있는 것으로 평가되었다.

III. 연구설계

1. 연구모형 및 변수설계

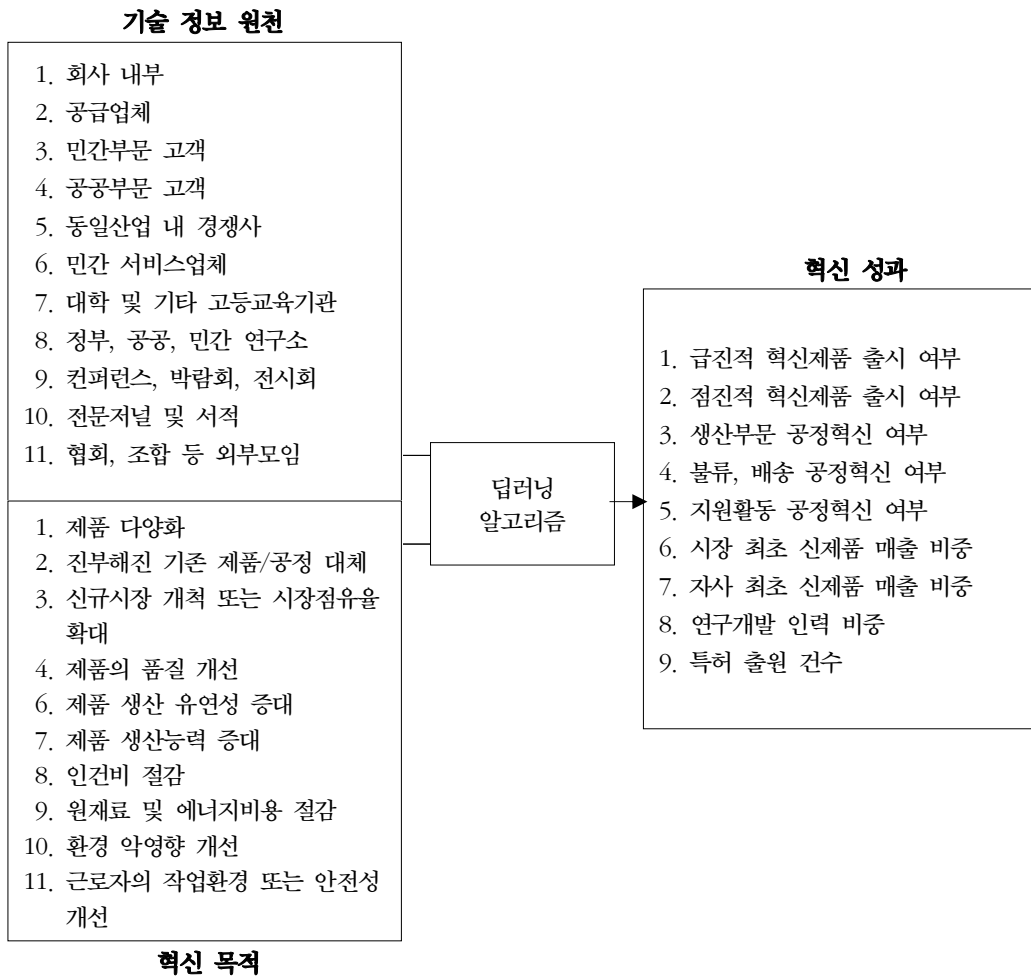
본 연구에서는 과학기술정책연구원(STPEI)이 실시하는 ‘한국기업혁신조사(KIS)’의 2016년 제조업 데이터를 활용하였다.²⁾ 한국기업혁신조사는 OECD Oslo Manual에 기초하여 실시되며, 조사 모집단은 2016년 이전 3년간 활동한 기업 중 상시종사자 수 10인 이상의 제조업체이다. 이를 통해 얻어진 4000개 기업의 데이터 중 본 연구에서는 분석 대상 요인에 있어 결측치가 많은 자료와 이상치(outlier)를 제외한 2079개 데이터를 활용하여 분석을 진행하였다.

앞서 살펴본 Wang and Chein(2006)에서는 혁신성과를 예측하기 위한 투입(input) 요인으로 기술 정보의 원천 활용도와 혁신 목적을 활용하였으며 산출(output) 요인으로는 혁신성과와 관련된 요인들인 신제품 및 공정의 매출 비중, 특허의 수 등을 편집하여 사용하였다. 또한

2) 익명의 심사자가 지적한 ‘기업혁신조사 데이터는 빅데이터인가’ 하는 문제에 대해 연구진은 기업혁신조사 데이터가 흔히 3V(Volume, Variety, Velocity)로 표현되는 빅데이터의 조건을 모두 충족하지는 못하지만 다음의 이유로 방법론을 적용하는 데에는 무리가 없다고 판단하였다. 첫째, 빅데이터와 관련하여 비정형 데이터의 처리 및 활용에 대한 관심이 대두되고 있으나 정형 데이터의 처리 및 활용 역시 중요한 분야이다. 기업혁신조사는 3년주기로 조사되고 있는데, 2002년 최초 조사 기업 수는 약 3000여 개였으나 최근인 2016년 기업혁신조사의 데이터는 8000여 개로 그 수가 점점 늘어나고 있으며, 이렇게 조사된 데이터는 축적되고, OECD 매뉴얼에 따라 조사되므로 타국의 조사 데이터와도 연계할 수 있어 빅데이터의 영역으로 충분히 확장될 수 있다. 따라서 그 일부에 방법론을 적용해보는 것은 의미가 있다고 판단하였다. 둘째, 딥러닝의 경우 빅 데이터 분석에 활발히 사용되고 있는 방법론이지만 그 근간에는 퍼셉트론과 인공신경망이 있는데, 이는 1957년부터 연구되어 오던 방법론으로 빅데이터의 개념이 등장하기 이전부터 사용되던 방법론이다. 따라서 본 연구에 사용한 데이터가 빅 데이터의 일반적인 정의에서는 다소 벗어나더라도 방법론의 적용에는 무리가 없다고 판단하였다.

각 요인들은 OECD Oslo Manual에 기초하여 조사되었으므로 본 연구에서 활용한 KIS 2016 데이터와도 유사한 점이 많다.

본 연구의 주된 목적은 혁신 예측 방법론으로써 딥러닝 모형의 적용 가능성을 탐색하는 데에 있다. 따라서 Wang and Chein(2006)이 제시한 모형을 활용하는 방향으로 연구를 진행하였다. 이를 참고하여 본 연구에서 혁신성과 예측을 위해 사용한 모형은 다음 (그림 6)과 같다.



(그림 6) 연구모형

본 모형에서 투입요인은 기술 정보 원천 관련 11개 변수와 혁신 목적 관련 10개 변수로 설정하였으며, 각 변수들은 0, 1, 2, 3의 4점척도로 조사되었다. 다만 Wang and Chein(2006)은 신경망에 한번에 많은 변수가 들어갈 경우 다중공선성의 문제가 생길 수 있음을 지적하였는데,

이를 피하기 위해 요인분석을 거쳐 투입 변수의 수를 감소시키는 방법을 사용하였다. 본 연구에서도 투입 변수들에 대한 요인분석을 실시하였다.

산출요인인 혁신성과의 경우에는 투입요인에 비해 더 많은 전처리 과정을 거쳐야 했는데, 투입 요인들의 경우 일괄적으로 4점척도에 따라 조사된 반면 혁신 성과는 0과 1로 이루어진 더미 변수와 비율 척도가 혼재되어 있으며 비율척도의 경우에도 0과 1 사이의 값인 비중과 자연수인 건수가 혼재되어 있어 이러한 차이를 완화해야 했기 때문이다. 각 변수들에 대한 자세한 내용은 다음 <표 1>과 같다.

<표 1> 변수의 조작적 정의

변수	변수의 형태
급진적 혁신제품 출시 여부	출시하였음 = 1 출시하지 않았음 = 0
점진적 혁신제품 출시 여부	출시하였음 = 1 출시하지 않았음 = 0
생산부문 공정혁신 여부	공정혁신이 있었음 = 1 공정혁신이 없었음 = 0
물류, 배송 공정혁신 여부	공정혁신이 있었음 = 1 공정혁신이 없었음 = 0
지원활동 공정혁신 여부	공정혁신이 있었음 = 1 공정혁신이 없었음 = 0
시장 최초 신제품 매출 비중	비중을 소수로 입력
자사 최초 신제품 매출 비중	비중을 소수로 입력
연구개발 인력 비중	비중을 소수로 입력
특허 출원 건수	건수를 그대로 입력

이에 산출 요인은 변수를 하나로 합치는 작업을 실시하였는데 가장 차이가 큰 특허 출원 건수의 경우 min max scaling을 통해 0과 1 사이의 범위로 정규화하였으며, 변수들의 평균을 사용하였다.

IV. 실증분석 결과

1. 변수의 기술 통계량

본 연구에서 사용한 변수들의 기술 통계량은 다음 <표 2>와 같다.

〈표 2〉 변수의 기술통계량

구분	변수	N	평균	표준편차
기술 정보 원천	회사 내부	2079	2.52	0.75
	공급업체	2079	2.32	0.86
	민간부문 고객	2079	2.45	0.83
	공공부문 고객	2079	1.83	1.06
	동일산업 내 경쟁사	2079	2.19	0.88
	민간 서비스업체	2079	1.52	0.95
	대학 및 기타 고등교육기관	2079	1.31	0.97
	정부, 공공, 민간 연구소	2079	1.48	0.99
	컨퍼런스, 박람회, 전시회	2079	1.55	0.97
	전문저널 및 서적	2079	1.37	0.93
협회, 조합 등 외부모임	2079	1.35	0.91	
혁신 목적	제품 다양화	2079	2.50	0.73
	진부해진 기존 제품/공정 대체	2079	2.58	0.69
	신규시장 개척 또는 시장점유율 확대	2079	2.57	0.71
	제품의 품질 개선	2079	2.75	0.55
	제품 생산 유연성 증대	2079	2.45	0.73
	제품 생산능력 증대	2079	2.30	0.77
	인건비 절감	2079	2.51	0.76
	원재료 및 에너지비용 절감	2079	2.89	0.78
혁신 성과	환경 악영향 개선	2079	1.97	0.79
	근로자의 작업환경 또는 안전성 개선	2079	2.01	0.79
	급진적 혁신제품 출시 여부	2079	0.18	0.39
	점진적 혁신제품 출시 여부	2079	0.55	0.50
	생산부문 공정혁신 여부	2079	0.33	0.47
	불류, 배송 공정혁신 여부	2079	0.23	0.42
	지원활동 공정혁신 여부	2079	0.16	0.36
	시장 최초 신제품 매출 비중	2079	0.04	0.14
	자사 최초 신제품 매출 비중	2079	0.13	0.22
	연구개발 인력 비중	2079	0.11	0.11
특허 출원 건수	2079	1.93	6.53	
rescaling된 특허 출원 건수	2079	0.01	0.05	
혁신 성과 변수	2079	0.19	0.14	

기술 정보 원천의 중요도와 혁신 목적 중요도와 관련된 변수들은 모두 4점척도로 조사되었다. 급진적 혁신제품 출시 여부와 점진적 혁신제품 출시 여부는 0과 1로 조사되었다. 이 때 점진적 혁신제품 출시 여부의 평균값이 점진적 혁신제품 출시 여부의 평균값보다 높은 것으로

보다 급진적 혁신제품보다 점진적 혁신제품이 시장에 더 많이 출시되고 있음을 확인할 수 있다. 공정혁신의 경우 각각 제조, 물류, 지원 관련 공정혁신으로 나누어 조사되었으며 이 역시 0과 1로 조사되었다. 이외 시장 최초 신제품 매출 비중과 자사 최초 신제품 매출 비중의 경우 0과 1 사이의 비율로 측정되었으며 특히 출원건수의 경우 앞서 기술한 바와 같이 0과 1 사이로 정규화하여 사용하였다. 혁신 성과 변수는 혁신 성과에 해당하는 변수들의 평균으로, 학습에 직접적으로 사용되었다.

2. 요인분석

신경망에 변수를 투입하기에 앞서 투입변수의 수를 줄이기 위한 요인분석을 실시하였다. 아이겐값 1.0 이상인 요인은 세 개로 나타나 변수들을 세 개의 요인으로 분류하였다. 이 때 요인 부하량이 한 가지 요인에만 뚜렷이 높게 나오는 변수를 취하였는데, 이는 요인부하량이 한 요인에 0.6 이상, 나머지 요인에서 0.3 이하일 것을 기준으로 하였다. 요인 분석 결과는 다음 <표 3>과 같이 나타났다.

<표 3> 요인분석 결과

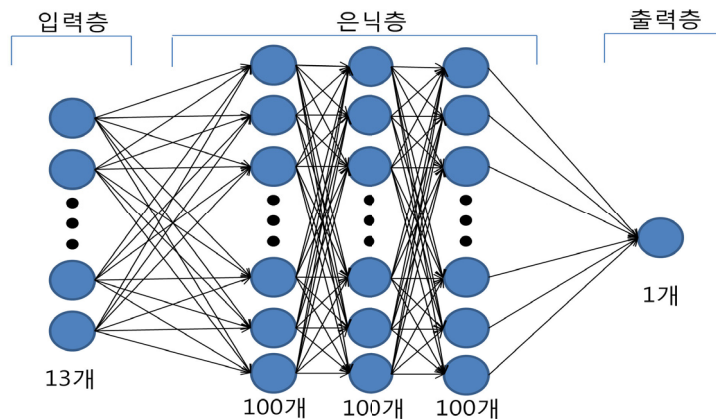
Variable	Factor1	Factor2	Factor3
민간 서비스업체	0.6399	0.1964	0.1126
대학 및 기타 고등교육기관	0.6855	0.1496	0.0993
정부, 공공, 민간 연구소	0.7071	0.1703	0.0748
컨퍼런스, 박람회, 전시회	0.7648	0.1641	0.1219
전문저널 및 서적	0.7675	0.0835	0.0334
협회, 조합 등 외부모임	0.6811	0.1215	-0.0929
제품 다양화	0.0633	0.1749	0.7386
진부해진 기존 제품/공정 대체	0.0582	0.2495	0.7109
신규시장 개척 또는 시장점유율 확대	0.0823	0.241	0.6859
제품의 품질 개선	0.1397	0.7336	0.2048
인건비 절감	0.1269	0.791	0.1815
원재료 및 에너지비용 절감	0.1526	0.8703	0.1359
환경 악영향 개선	0.1068	0.8606	0.1059
아이겐 값	1.1016	2.9515	1.6788
설명량(%)	0.4074	0.3876	0.2205
누적 설명량	0.4074	0.7950	1.0155
Cronbach's alpha	0.8651	0.9086	0.8204

요인분석 결과 투입 요인은 크게 세 가지 요인으로 묶을 수 있었다. 첫 번째 요인에 해당하는 변수들은 정보의 원천 중 민간 서비스업체, 대학 및 기타 고등교육기관, 정부, 공공, 민간 연구소, 컨퍼런스, 박람회, 전시회, 전문저널 및 서적, 협회, 조합 등 외부모임에 해당하는 변수들이었으며, 나머지 변수들은 유의미한 요인부하량을 가지지 못하여 삭제하였다.

두 번째 요인에 해당하는 변수들은 혁신 목적 중 인건비 절감, 원재료 및 에너지비용 절감, 환경 악영향 개선, 근로자의 작업환경 또는 안전성 개선이었으며 세 번째 요인은 혁신 목적 중 제품 다양화, 진부해진 기존 제품/공정 대체, 신규시장 개척 또는 시장점유율 확대였다. 나머지 변수들은 유의미한 요인부하량을 가지지 못하여 삭제하였다. 세 요인의 Cronbach's alpha 값은 모두 0.8 이상으로 높게 나타나 유의미한 요인으로 해석할 수 있었다.

3. 딥러닝을 사용한 예측 모형

변수들을 투입하여 혁신성과를 예측하기 위한 모델의 뉴런은 다음 (그림 7)와 같이 구성되었다. 모델 구축에는 Python 3.60 아나콘다 배포판과 Tensorflow 1.3 버전이 사용되었다.



(그림 7) 뉴런의 구성

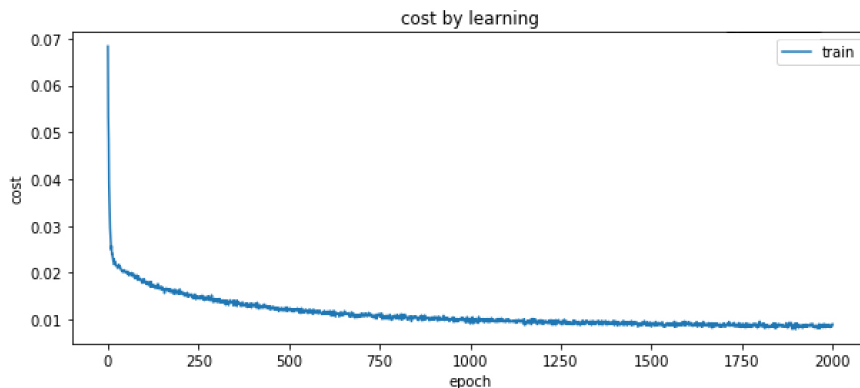
입력층의 경우 13개로 구성되었는데, 이는 요인분석 결과 유의한 부하량을 가진 변수가 13개이기 때문이다. 그리고 출력층은 하나로 구성하였는데, 예측하고자 하는 혁신 성과를 하나의 변수로 만들었기 때문이다. 그리고 은닉층은 총 세 개로 구성하였으며 각각 100개의 뉴런을 배치하여 신경망을 구현하였다.

데이터는 학습 데이터(training data)와 테스트 데이터(test data)로 구분하여 학습을 통한 모델 구축에는 학습 데이터를, 예측 정확도 파악에는 테스트 데이터를 사용하였다. 이 때 학습 데이터는 전체 데이터의 80%, 테스트 데이터는 전체 데이터의 20%로 설정하였다. 이 때 데이터의 분할은 무작위 추출을 통해 진행하였으며, 전체 데이터인 2079개의 약 80%에 해당하는 1663개를 추출하여 학습 데이터로, 나머지 416개 데이터를 테스트 데이터로 설정하였다. 학습에는 Adam Optimizer를 사용하였으며 학습과정에는 과적합(overfitting)을 피하기 위해 30%의 drop out을 적용하였다. 본 모델은 오차를 최소화하는 방향으로 가중치를 변화시키며 학습해 나가는데, 오차로는 평균제곱오차(mean squared error)를 활용하였다.

4. 분석 결과

1) 학습 및 예측 결과

학습을 위해 학습 데이터를 1000회 학습시켰으며, 그 결과 오차는 다음 (그림 8)과 같이 줄어들고 있음을 확인할 수 있었다.



(그림 8) 학습에 따른 오차 감소

학습 결과 오차를 0.01 수준으로 감소시킬 수 있었다. 학습한 모델을 시험 데이터에 적용하였을 경우 오차는 0.02 수준으로 나타났는데, 앞서 Wang and Chein(2006)의 연구에서 오차가 0.04~2.44의 범위에 분포하였던 것과 비교하여 더 높은 예측력을 가짐을 확인할 수 있었다.

다만 본 연구에서 학습 데이터와 테스트 데이터의 구분은 무작위 추출 방식으로 진행되었는

데, 이는 추출된 데이터에 따라 오차가 달라질 수 있음을 의미한다. 따라서 추출과 학습을 여러 차례 진행하여 비슷한 수준의 오차를 보이는지 확인할 필요가 있다. 이에 추출 및 학습을 5회 진행하였으며, 그 결과는 다음 <표 4>와 같이 나타났다.

<표 4> 5회 반복 시행 결과

횟수	학습 오차	테스트 오차
1	0.009	0.021
2	0.008	0.021
3	0.008	0.022
4	0.009	0.023
5	0.008	0.020
평균	0.0084	0.0214

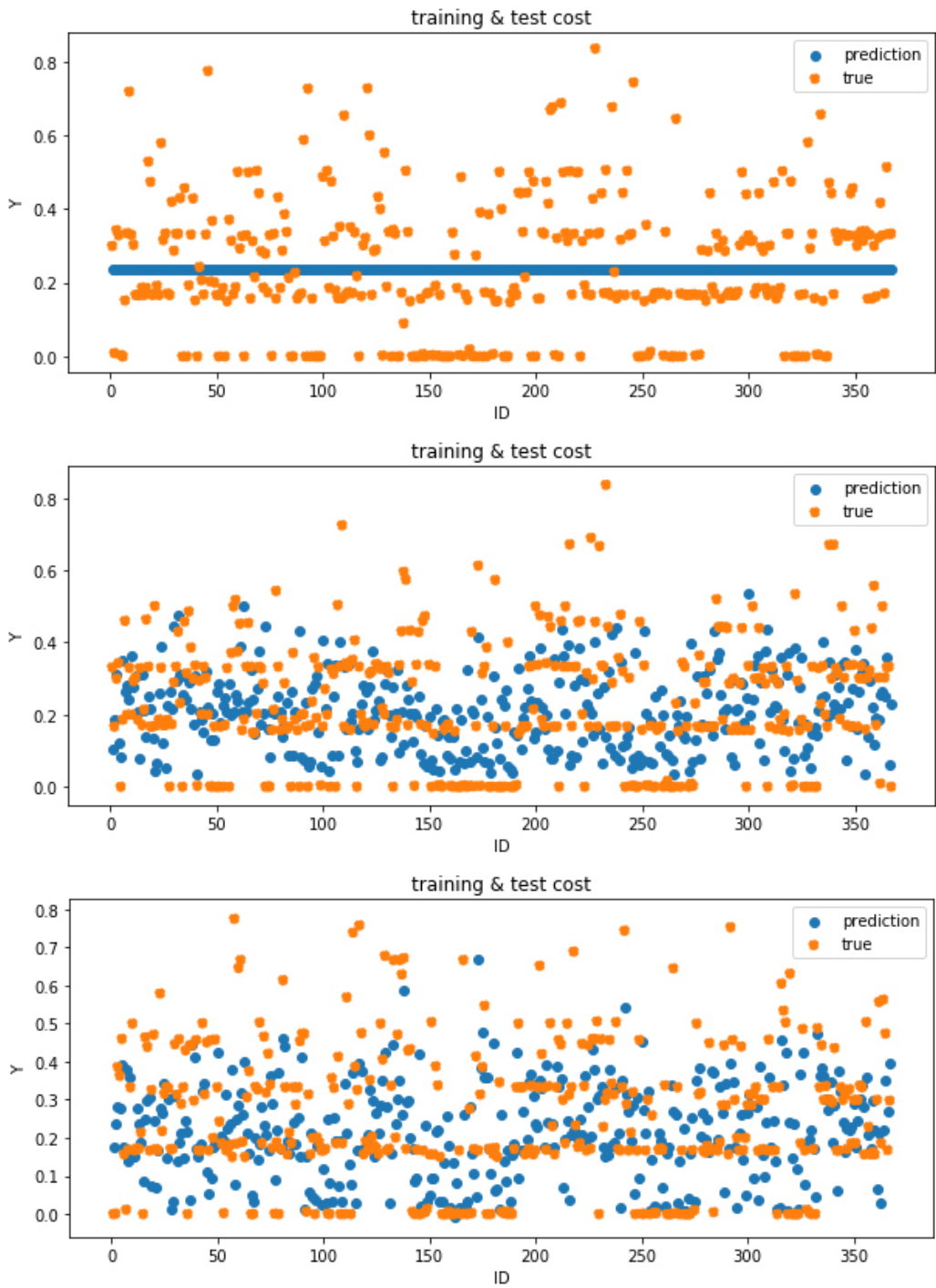
학습 오차는 모두 0.01 미만을 보였으며 테스트 오차는 0.021~0.023 수준인 것으로 나타났다. 이에 추출에 따른 오차의 수준은 크게 달라지지 않는 것을 알 수 있었으며, 5회 반복시행의 결과 모두 Wang and Chein(2006)의 연구결과에 비해 낮은 것으로 나타났다.³⁾

또한 학습이 진행됨에 따라 예측의 자유도가 달라지는 것을 알 수 있었는데, 이는 (그림 9)와 같다.

(그림 9)에서 파란색 점은 딥러닝 모형을 통해 예측한 점을 나타내며 주황색 점은 실제 데이터를 나타낸다. 가장 위에 있는 그래프는 학습이 제대로 이루어지지 않았을 때의 예측 양상이고 가운데는 어느 정도 학습이 이루어졌을 때의 예측 양상이며 마지막 그래프는 학습이 가장 잘 이루어졌을 때의 예측 양상이다. 학습이 잘 이루어지지 않았을 경우 모든 예측이 같은 값을 가리키는 것과 대비하여 학습이 이루어질수록 중심에서 먼 값까지도 예측을 해내는 것을 확인할 수 있었다.

혁신 성과와 관련된 선행연구들은 모두 회귀분석 모형을 활용하였고, 이는 혁신 성과에 영향을 미치는 변수들을 설명할 때에는 유의미한 결과를 제시하였으나 예측을 시도한 경우는 거의 존재하지 않았다. 또한 인공지능망 모형을 사용한 Wang and Chein(2006)의 연구결과와 비교하였을 때에도 더 적은 예측 오차를 보임을 확인하였다. 이에 혁신 연구에 있어 딥러닝 모형의 적용 및 향후 발전 가능성을 확인할 수 있었다.

3) 익명의 심사위원이 지적한 기존 방법론과의 예측력 비교 분석의 경우 동일 데이터에 단순회귀분석을 적용하였을 때 모델 전체의 오차가 0.03 수준으로 나타나 딥러닝 모형의 예측력보다는 다소 높은 오차를 보이는 것을 확인하였으나 모델의 설명력에서 나오는 오차와 예측 오차는 동일선상에서 비교하기에 무리가 있다고 판단하여 후속연구로서 보다 엄밀한 비교 연구를 진행하기로 하였다.



(그림 9) 학습에 따른 예측 변화

V. 결론 및 시사점

본 연구에서는 빅데이터 분석방법을 활용하여 기업의 혁신 성과를 예측하는 알고리즘을 제시하고자 하였고 그 중 딥 러닝을 활용한 알고리즘을 통해 이를 달성하고자 하였다. 이에 딥 러닝 모델에 KIS 2016 데이터를 학습시켜 예측모델을 구축하였다.

분석 결과 예측모형의 학습 오차는 0.008 수준으로 학습되었고 예측 오차는 0.02 수준임을 확인할 수 있었다. 또한 학습이 진행됨에 따라 예측의 자유도도 더 높아짐을 확인할 수 있었는데, 학습이 제대로 되지 않은 모형을 활용할 경우 예측값이 중간지점에 있는 하나의 값으로 모아지는 반면 학습이 진행될수록 중간에서 먼 값까지도 예측이 가능해짐을 확인할 수 있었다. 선행연구들과 비교하였을 때 본 연구는 혁신 연구에 딥 러닝 모형의 적용가능성을 확인하였다는 의의를 가지며, Wang and Chein(2006)과 비교하여 더 낮은 수준의 오차를 보여 모델의 발전 가능성을 확인하였다는 의의도 지닌다고 할 수 있다.

또한 Wang and Chein(2006)의 연구에 비해 방법론적인 부분에 있어 보다 진보된 방법론을 적용할 수 있었다. 우선 데이터의 수에 있어 Wang and Chein(2006)은 53개의 기업 데이터를 활용한 반면 본 연구에서는 2079개 기업의 데이터를 활용하였다. 여기에 보다 복잡한 모형을 적용할 수 있었는데, 이는 최근 고안된 ReLU 함수나 AdamOptimizer 등을 사용함으로써 보다 효과적인 학습이 가능하였기 때문이다. 기존 신경망에 사용하던 sigmoid 함수나 Wang and Chein(2006)이 사용한 hyperbolic tangent 함수에 비해 본 연구에서 사용한 ReLU 함수는 gradient vanishing 문제를 어느 정도 해결할 수 있었는데, 이를 통해 은닉층을 깊게 할 수 있었다.

하지만 본 연구는 다음과 같은 한계점도 가진다. 빅 데이터 분석 방법의 경우 빅데이터가 방대하고, 빠르게 축적되며 비정형성을 지닌다는 특성 때문에 데이터의 전처리를 생략할 수 있는 알고리즘을 필요로 하나 본 연구에서는 투입 요인과 산출 요인에 있어 둘 다 전처리 이후 모델에 적용하였다. 이 부분에 있어서는 모델의 지속적인 개량이 필요할 것이다.

이러한 한계점에도 불구하고 본 연구는 혁신 예측 분야에 잘 적용되지 않던 빅데이터 및 인공지능을 접목시키고자 했다는 데에 의의가 있다. 혁신과 기술에 대한 예측은 본질적으로 어려운 분야이지만, 이러한 시도가 지속된다면 어려움을 경감하는 데에 상당한 도움을 줄 수 있을 것으로 기대한다. 또한 향후 후속연구로서 기존에 사용되었던 회귀분석 모형과 딥 러닝 모형의 효율성 및 예측력을 비교하는 연구도 추진할 수 있을 것이다.⁴⁾

4) 익명의 심사자가 지적한 본 논문의 방법론과 선행연구에서 사용한 방법론간의 예측력 차이의 경우 중요한 연구주제이나 본 논문의 주제인 '혁신연구 분야에 딥 러닝 알고리즘의 적용가능성 탐색'을 다소 벗어나는 큰 주제이기 때문에 향후 이에 대한 연구를 후속연구로서 진행하기로 한다.

참고문헌

- 김성홍·김진한 (2011), “개방형 제품혁신 성과에 관한 탐색적 연구 : 제품혁신 전략, 원천, 파트너 기여도 관점”, 「대한경영학회지」, 24(2): 685-703.
- 김인중·나기현·양소희·장재민·김운중·신원영·김덕중 (2010), “딥러닝과 통계 모델을 이용한 T-커머스 매출 예측”, 「정보과학회논문지」, 44(8): 803-812.
- 사이토 고키 (2017), 「밑바닥부터 시작하는 딥러닝」, 서울 : 한빛미디어.
- 성태경 (2005), “기업의 기술혁신성과 결정요인 : 기업규모와 외부 네트워크의 역할을 중심으로”, 「대한경영학회지」, 18(4): 1767-1788.
- 이우식 (2017), “딥러닝분석과 기술적 분석 지표를 이용한 한국 코스피주가 지수 방향성 예측”, 「한국데이터정보과학회지」, 28(2): 287-295.
- 이우식·전희주 (2016), “딥러닝 분석을 이용한 중국 역내·외 위안화 변동성 예측”, 「한국데이터정보과학회지」, 27(2): 327-335.
- 이지훈 (2016), “딥러닝을 이용한 주가 예측 모델”, 숭실대학교 대학원 석사학위 논문.
- 정선양 (2016), 「전략적 기술경영」, 제4판, 박영사.
- 정한웅 (2016), “딥러닝 알고리즘에 기반한 기업부도 예측”, 한양대학교 대학원 석사학위 논문.
- 천종기·신용준·배후석 (2010), “기술혁신역량과 SCM활동이 기업혁신성과에 미치는 영향: 신발산업을 중심으로”, 「산업혁신연구」, 27(2): 25-57.
- Hinton, G. E. (2006), “Training Products of Experts by Minimizing Contrastive Divergence”, *Neural Computation*, 14(8): 1771-1800.
- Hopfield, J. J. (1982), “Neural Networks and Physical Systems with Emergent Collective Computational Abilities”, In *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, 411-415.
- Minsky, M. and Papert, S. (1969), *Perceptrons*, Cambridge, MA: MIT Press.
- Tidd, J. and Bessant, J. (2013), *Managing Innovation: Integrating Technological, Market and Organizational Change*, 5th Edition, Chichester: Wiley.
- Wang, T. Y. and Chien, S. C. (2006), “Forecasting Innovation Performance via Neural Networks: A Case of Taiwanese Manufacturing Industry”, *Technovation*, 26(5): 635-643.

황정재

건국대학교 밀러MOT스쿨에서 기술혁신경영 전공으로 석사학위를 취득하였으며 현재 한국과학기술기획평가원(KISTEP)에 재직중이다. 관심연구분야는 혁신체제론, 지역혁신, AI, 딥러닝 등이다.

김재영

건국대학교 밀러MOT스쿨에서 국가기술경영 전공으로 석사학위를 취득 후 동 대학원 박사과정에 재학중이다. 관심연구분야는 국가혁신체제, 적정기술, R&D 성과분석 등이다.

박재민

미 오하이오주립대학교에서 경제학 박사학위를 취득하고 현재 건국대학교에서 기술경영학과 교수로 재직 중이다. 관심연구분야는 기술트렌드, 가치경영, R&D 성과분석 등이다.